Collaborative Filtering (CF)

- Data Source: Primarily focuses on user interaction data. In our case, it leverages playtime metrics, sentiment scores derived from reviews, and engagement metrics such as votes and the number of games owned. This approach does not require information about the games themselves (like genre or developer) to make recommendations.
- Methodology: Uses the Top-K Nearest Neighbors algorithm to find similarities among users based on their interactions and preferences. Recommendations are made based on the preferences of similar users, emphasizing a "user-to-user" matching process. The algorithm computes similarity scores using a vector space model and cosine similarity measure to identify users with similar tastes.

Content-Based Filtering (CBF)

- Data Source: Utilizes detailed attributes of items (in this case, games) such as developer, publisher, genres, tags, and descriptions from the Steam App Info dataset, alongside user-generated content (reviews) for a more nuanced understanding of game content.
- Methodology: Constructs feature vectors from game characteristics and user profiles, then applies cosine similarity to compare these vectors. Recommendations are generated based on the similarity between a user's profile and game attributes, focusing on aligning game characteristics with user preferences. This "item-to-user" matching emphasizes the content of the items themselves rather than user interaction patterns.

Key Differences

- Basis of Recommendations: CF is based on the idea that users will like items similar to other items they have liked, or items liked by similar users. CBF recommends items by matching item features with user preferences.
- Data Used: CF relies on user behavior (e.g., playtime, review sentiment, engagement metrics), while CBF focuses on item attributes (e.g., tags, descriptions, genres) and textual analysis of reviews.
- Algorithms: The methodologies and algorithms applied in CF (Top-K Nearest Neighbors for user similarity) and CBF (cosine similarity for matching user profiles to item attributes) are fundamentally different.